A Novel Way To Automatically Mine Comparable Entities From Comparative Questions

P.Ujwala Sai, 2 N.Rajesh Babu
1,2Dept. of CSE, PYDAH College of Engineering, Patavala,Kakinada, E.g.dt,AP, India

ABSTRACT:
A comparison activity usually contains search for applicable web pages containing information about the targeted products, find competing products, read reviews and identify pros and cons. In this paper we focus on finding a set of comparable entities given a user’s input entity. Comparing one thing with another is an archetypal part of human decision making process. Conversely it is not all the time simple to know what to compare and what are the substitutes. To deal with this complexity we present a novel way to automatically mine comparable entities from comparative questions that users posted online. To make sure high accuracy and high recall we develop a weakly-supervised bootstrapping method for comparative question recognition and similar entity extraction by leveraging a large online question archive. Both considerably do better than an existing state-of-the-art method.

RELATED WORK:
To the preeminent acquaintance this is the first effort to particularly address the difficulty on finding good comparators to hold up user comparison activity. To suggest using comparative questions posted online that replicate what users truly care about as the medium from which we mine comparable entities. In terms of determining related items for an entity our work is alike to the research on recommender systems which recommend items to a user. Recommender systems mainly depend on similarities between items or their statistical correspondences in user log data. Bootstrapping process has been shown to be very effectual in previous information extraction research. Our work is similar to them in terms of line of attack using bootstrapping technique to extract entities with a specific relation. Though the task is different from theirs in that it requires not only extracting entities, comparator extraction but also make certain that the entities are extracted from comparative questions.

EXISTING METHOD:
Comparator mining is associated to the research on entity and relation extraction in information extraction. Particularly the most applicable work is mining comparative sentences and relations. Their methods applied class sequential rules (CSR) and label sequential rules (LSR) educated from annotated corpora to recognize comparative sentences and take out comparative relations correspondingly in the news and review domains. The same methods can be applied to comparative question identification and comparator mining from questions.

DISADVANTAGES:
These methods characteristically can attain elevated accuracy but undergo from low recall.

**PROPOSED METHOD:**

The proposal of a novel weakly supervised method to make out comparative questions and extract comparator pairs concurrently. We depend on the key insight that a good comparative question classification pattern should extract good comparators and a good comparator pair should occur in good comparative questions to bootstrap the withdrawal and classification procedure. By leveraging huge amount of unlabeled data and the bootstrapping procedure with slight supervision to conclude four parameters. CliqueGrow is an agglomerative algorithm that aims to group entities into clusters of mutually comparable entities. Once clusters are identified, any two entities belonging to the same cluster are comparable.

**ADVANTAGES:**

To make certain high precision and high recall we develop a weakly-supervised bootstrapping method for comparative question identification and comparable entity extraction by leveraging a large online question archive.

**SYSTEM ARCHITECTURE:**

**ADMIN:**

a) Add Dataset: after admin login he can add comparable entity questions related to mobile, laptop and bike into database.

b) Comparable entity questions mining: here we are separating comparable questions and non-comparable questions.

c) Clustering comparable entities: CliqueGrow is an agglomerative algorithm that aims to group nodes into clusters of mutually comparable entities. Once clusters are identified any two nodes belonging to the same cluster are comparable.

d) View Result: here displaying recall and precision values of proposed clustering method comparing with existing method.

**USER:**

a) Search query: here user issuing comparable entity questions along with displaying result information.

B) View History: displaying user search history information

**CLIQUEGROW CLUSTERING ALGORITHM:**

CliqueGrow starts with seed unit clusters and iteratively merges other base structures, until they con verge to original clusters. We use triangles as initial seeds because a triangle is the basic unit of transitive closure that is observed. A triangle defines a unique topic among the three pairs of comparable entities of a triangle. Using triangles as seeds, we gradually grow clusters, by connecting to neighboring entities.

By the nature of an agglomerative approach, the topic purity is diluted as the cluster grows. We thus quantify the quality of triangles and populate a priority queue H, to expand only high quality triangles. The quality is quantified as the lowest edge weight of a triangle, as a triangle with a high quality score corresponds to the clique in which every pair co-occurs frequently.

**ALGORITHM STEPS:**

1. Retrieve all triangle structures from the dataset and insert them into a priority queue H, ordered by the minimum of the edge weights.
2. Pick the top seed in the sorted queue, and compute its probability.
3. Compute for each neighboring base structure B from seed. Include the corresponding structure in the cluster, if CP(S,B) > clustering threshold (CT).
4. Update the probability for the expanded cluster at this iteration.
5. Go to Step 3 and repeat until expansion ceases.
6. Remove the clustered triangles from queue go to Step 1 and iterate until queue is empty
EXPERIMENTAL RESULTS:

<table>
<thead>
<tr>
<th>Channel</th>
<th>Gap</th>
<th>Polo</th>
<th>Kobe</th>
<th>Canon</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dior</td>
<td>Nike</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Louis Vuitton</td>
<td>Tommy Hilfiger</td>
<td>Dior</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Coach</td>
<td>PUMA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Geox</td>
<td>Geox</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Prada</td>
<td>DUCANOEVO</td>
<td>Prada</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Esurance</td>
<td>Old Navy</td>
<td>Fender</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Victoria’s Secret</td>
<td>Old Navy</td>
<td>Fender</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>JVC</td>
<td>JVC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Nike</td>
<td>Nike</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Giani</td>
<td>Giani</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Examples of comparators for different entities:
As shown in the table, the comparator mining method productively discovers realistic comparators. For example for „Chanel”, most results are high-end fashion brands such as „Dior” or „Louis Vuitton” while the ranking results for „Gap” typically enclose similar apparel brands for young people such as „Old Navy” or „Banana Republic”. For the basketball player „Kobe”, most of the top ranked comparators are also famous basketball players. Some interesting comparators are shown for „Canon” the company name. It is well-known for different kinds of its products for example, digital cameras and printers so it can be compared to different kinds of companies.

CONCLUSION:
We replacing a weakly supervised bootstrapping with clique grow clustering technique to identify comparative questions and take out comparable entities at the same time. Existing weakly supervised indicative extraction pattern mining method is a pattern-based approach but it is dissimilar in a lot of aspects such as an alternative of using various class sequential rules and label sequential rules, our process aims to become skilled at Clustering technique which can be able to be used to identify comparative questions and take out comparators concurrently.

REFERENCES:

**Miss.P.Ujwala Sai** is a student of PYDAH College of Engineering, Patavala. Presently she is pursuing her M.Tech [Computer Science and Engineering] from this college and she received her B.Tech from Sri Sai Aditya Institute of Sciences and Technology, affiliated to JNT University, Kakinada in the year 2009. Her area of interest includes Data Mining and Data warehousing and Object oriented Programming languages.

**Mr.N.RajeshBabu**, well known Author and excellent teacher received M.Tech (CSE) from Acharya Nagarjuna university, is working as Associate Professor and HOD, Department of B.Tech, M.Tech Computer science engineering , Pydah college of Engineering , He is having rich experience in teaching subjects of Electronics and Communication Engineering and Computer Science and Engineering . He has published many technical papers in National and International journals .His area of Interest includes Programming Languages,Data mining and Data warehousing,Databases and Computer Neworks.